Routledge
Taylor & Francis Group

Journal of Risk Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjrr20

The effectiveness of risk assessments in risk workshops: the role of calculative cultures

Lucia Bellora-Bienengräber, Clemens Harten & Matthias Meyer

To cite this article: Lucia Bellora-Bienengräber, Clemens Harten & Matthias Meyer (2023) The effectiveness of risk assessments in risk workshops: the role of calculative cultures, Journal of Risk Research, 26:2, 163-183, DOI: 10.1080/13669877.2022.2108120

To link to this article: https://doi.org/10.1080/13669877.2022.2108120

JOURNAL OF RISK RESEARCH
2023, VOL. 26, NO. 2, 163–183
https://doi.org/10.1080/13669877.2022.2108120

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

The effectiveness of risk assessments in risk workshops: the role of calculative cultures

Lucia Bellora-Bienengräber[a], Clemens Harten[b] and Matthias Meyer[b]

[a]Department of Accounting, University of Groningen, Groningen, Netherlands; [b]Institute of Management Accounting and Simulation, Hamburg University of Technology, Hamburg, Germany

ABSTRACT

This paper investigates drivers of the effectiveness of risk assessments in risk workshops dominated by 'quantitative skepticism'. Moreover, it contrasts our findings with those of previous research that assumed the dominance of 'quantitative enthusiasm'. Quantitative skepticism is a calculative culture characterized by an attitude that regards risk assessments as learning tools supporting the holistic formation of judgments incorporating difficult-to-quantify information. It contrasts with quantitative enthusiasm, which is a calculative culture that considers risk assessments as fully descriptive of reality. Prior research primarily focused on understanding the effectiveness of risk assessments under a calculative culture of quantitative enthusiasm. To understand what drives the correctness of risk assessment and the time needed to assess risks in workshops under a calculative culture of quantitative skepticism, we use an agent-based model that simulates risk assessment with risk workshops and that models agents' cognitive processes using ECHO, a constraint satisfaction network (CSN). Our simulations show that, compared to risk workshops under conditions of quantitative enthusiasm, there are often lengthy periods of stagnation in individual and collective risk assessments and a strong path dependency on discussions. Prioritizing concerned participants improves the correct assessment of high risks at the expense of the correct assessment of low risks. Notwithstanding similarities in the drivers of the effectiveness of risk assessment across different calculative cultures, our results show that the predominant calculative culture matters when—to enhance their effectiveness—designing and implementing risk workshops.

ARTICLE HISTORY
Received 3 February 2022
Accepted 12 July 2022

KEYWORDS
Calculative culture; enterprise risk management; risk assessment; risk workshop; agent-based modeling; ECHO network

Introduction

This paper investigates drivers of the effectiveness of risk assessments in risk workshops dominated by a calculative culture of 'quantitative skepticism'. Moreover, it contrasts our findings with those of previous research that assumed the dominance of 'quantitative enthusiasm'. A calculative culture captures 'attitudes towards the use and limitations of highly analytical calculative practices in an organization' (Mikes 2009, 21). Mikes (2009) distinguishes between two calculative cultures, i.e. quantitative enthusiasm and quantitative skepticism. Quantitative enthusiasm considers the risk assessments made in the organization as representations of a measurable

CONTACT Lucia Bellora-Bienengräber l.bellora-bienengraeber@rug.nl Faculty of Economics and Business, Department of Accounting, University of Groningen, Nettelbosje 2, Groningen, 9747 AE, The Netherlands
Supplemental data for this article is available online at https://doi.org/10.1080/13669877.2022.2108120.

economic reality (Mikes 2009, 35). Quantitative skepticism, by contrast, entails a calculative culture that regards risk assessments as learning tools supporting the holistic formation of judgments incorporating difficult-to-quantify information (Burchell et al. 1980; Mikes 2009; Power 2007). The resulting risk numbers are seen as trend indicators, rather than a full account of reality (Mikes 2009).[1]

We expect that an organization's predominant calculative culture[2] affects the unfolding of discussion processes for risk assessments and, thus, their outcomes. At present, the scant literature investigating the effectiveness of risk assessments in an organization's risk workshops does not explicitly address the effects of different calculative cultures and implicitly assumes a calculative culture of quantitative enthusiasm. To the best of our knowledge, this is the first paper that contrasts the drivers of the effectiveness of risk assessments, depending on the predominant calculative culture. Specifically, we expect that the design and implementation of risk workshops—intended to enhance their effectiveness—should be different depending on the predominant calculative culture.

A risk workshop is effective when it correctly assesses the considered risk, ideally minimizing the time needed to do so (van Asselt and Renn 2011; Harten, Meyer, and Bellora-Bienengräber 2022; Quail 2011). The correct assessment of risks[3] is a non-trivial task for organizations. The ability to distinguish between high and low risks is vital for any business. Nevertheless, such an assessment can be challenging as related information[4] is often distributed inside and outside the organization (Neef 2005). Risk workshops are a common tool to aggregate information about risks (COSO 2017). During risk workshops, participants discuss risks and derive an assessment of their impact and likelihood (Boholm and Corvellec 2016). To improve comparability with prior research (Harten, Meyer, and Bellora-Bienengräber 2022), we focus this study on the assessment of a risk's likelihood.[5]

Previous research has struggled to systematically disentangle different drivers of effectiveness in risk workshops. A notable exception is Harten, Meyer, and Bellora-Bienengräber (2022), who conceptualize risk workshops as transactive memory systems. Transactive memory systems combine the expertise of individuals by accessing their individual information in a discursive process (Wegner 1987). Thus, transactive memory systems include individual knowledge, knowledge about who knows what, and communication to access each other's knowledge. Within such a theoretical framework, a calculative culture determines how individuals process information to form risk assessments.

In Harten, Meyer, and Bellora-Bienengräber (2022), although not explicitly stated, the underlying conceptualization of information processing and formation of assessments is akin to what has been described above as quantitative enthusiasm. Indeed, in their model, risk workshop participants exchange information and gradually update their probabilistic beliefs about risks. This results in an overall quantitative risk assessment, which measures the overall likelihood of the discussed risk. This raises the question on the extent to which their results can be generalized when considering the boundary condition of the predominant calculative culture. In other words, would the results of Harten, Meyer, and Bellora-Bienengräber (2022) also hold under a calculative culture of quantitative skepticism? In the latter case, risk workshops provide an opportunity to exchange risk-relevant facts and their effects concerning an organization's risks, thereby improving the understanding of the overall risk environment. We expect that the difference in the predominant calculative culture is crucial for a better understanding of the effectiveness of risk assessments in risk workshops. For example, Harten, Meyer, and Bellora-Bienengräber (2022) observation that longer stagnation phases in discussions may indicate a correct assessment of risks might no longer be justified. In a calculative culture of quantitative skepticism, risk assessments might remain stable for a longer time, even when new information is presented. New information that is largely consistent with the participants' previous beliefs is absorbed instead of slightly changing the overall risk assessment.

To investigate our research question, we use agent-based modeling (ABM). We simulate how risks are assessed in risk workshops. The simulated agents are the participants who—without any hidden agenda (i.e. without own undisclosed objectives)—exchange information to reach a risk assessment for one specific risk. ABM provides a computational laboratory for controlled experiments where agents act according to predefined rules in a clearly defined environment (Wall and Leitner 2021). Harten, Meyer, and Bellora-Bienengräber (2022) already leverage this method to address the challenge of investigating cognitive processes and the absence of an objective benchmark for actual risk assessments (McNamara and Bromiley 1997). ABM allows to model the development of an individual's knowledge during the discussion and the group's utilization of individual knowledge to reach a risk assessment (Secchi 2015; Wall and Leitner 2021). Specifically, ABM permits the representation of diverse types of information processing and judgment formation of individuals, thereby incorporating different calculative cultures.

While Harten, Meyer, and Bellora-Bienengräber (2022) model agents' information processing and the formation of risk assessments using a Bayesian network, we use ECHO, a constraint satisfaction network (CSN) (Thagard 2012). Harten, Meyer, and Bellora-Bienengräber (2022) cognitive architecture of agents represents a calculative culture of quantitative enthusiasm, as Bayesian networks use probabilities as input and provide a probabilistic risk assessment as output, which can be interpreted as the measurement of the underlying overall risk. ECHO models, by contrast, represent a calculative culture of quantitative skepticism as they encode individual risk assessments as coherence-based relationships typifying qualitative mental models in the cognitive architecture of agents. Their output can be interpreted as a holistic judgment (vs. a probabilistic measurement of economic reality) that incorporates difficult-to-quantify information. To enable a direct comparison of the results, we follow Harten, Meyer, and Bellora-Bienengräber (2022) in our basic experimental design.[6]

Compared to a numerically generated benchmark, we investigate scenarios that are more realistic. We simulate the effects of (1) limits to information transfer within the group, (2) incomplete discussions involving various approaches to terminate the discussion and to make a decision, (3) group characteristics like information distribution and hierarchical relationships, and (4) rules specifying interaction patterns (e.g. prioritizing participants who are concerned about a specific risk). As we conduct these experiments with agents who have a calculative culture of quantitative skepticism, we can assess which of Harten, Meyer, and Bellora-Bienengräber (2022) results, rooted in a calculative culture of quantitative enthusiasm, still hold and which, in turn, are associated with the change of the calculative culture.

We find that the type of calculative culture predominant in an organization matters for the effectiveness of risk assessments in risk workshops. Indeed, some of the drivers of a risk assessment's effectiveness in risk workshops are different from those found by Harten, Meyer, and Bellora-Bienengräber (2022). Concerning the development of the discussion over time, we document an overall improvement of risk assessments for both high and low likelihood risks. However, we regularly observe sudden and seemingly unpredictable changes in participants' and groups' risk assessments when new information overturns the previously stable beliefs of the participants. The exchange of critical information can rapidly shift assessments, questioning the stagnation of a discussion as a criterion for ending it. At the same time, consensus serves as a good indicator of the risk assessment's correctness. We find that path dependencies, given quantitative skepticism, are characteristic of discussion processes, i.e. what matters for the risk assessment is not only *what* information is exchanged but also *when* it is exchanged. Initially, new information has the potential to overturn the participant's mental model, which contains little information at this early stage. Finally, the prioritization of concerned participants only results in the highest level of risk assessment correctness for high risks, while hierarchical differences among participants do not negatively affect the correct assessment of risks.

This paper makes at least three contributions to theory and practice. First, to the best of our knowledge, our paper is the first to show that different calculative cultures result in different drivers of risk assessments' effectiveness. The identification of the distinction between the calculative cultures of quantitative enthusiasm and quantitative skepticism, respectively, is important as a boundary condition when designing and implementing risk workshops aimed at the highest possible risk assessment correctness. The reason for this, among others, is that risk workshops in different organizations require different durations, different termination rules, and different treatments of hierarchical differences and dissenting participants.

Second, despite these differences, we also conclude that a few drivers of the effectiveness of risk assessments are resistant to changes in the dominant calculative culture. For example, the tradeoff between correctly identifying low and high risks as well as the negative effect of unequal information distribution within the group are independent of the calculative culture being quantitatively enthusiastic or skeptical. This points at the ubiquitous importance of considering these process characteristics and drivers when facilitating a risk workshop.

Third, we introduce a novel methodological and conceptual approach to risk literature that incorporates the way in which (i.e. *how*) risk information is processed and used for risk assessments in ECHO networks.

Theoretical background

Risk assessments in risk workshops

Risk workshops are used to assess the impact and likelihood of risks (Boholm and Corvellec 2016; COSO 2017). The workshops are discussions moderated by a facilitator and they enable a leader to decide using the outcome of a workshop. The group can use its participants' diverse backgrounds by aggregating their individual knowledge and thereby reach better decisions than the participants would have reached on their own (LiCalzi and Surucu 2012; Lu, Yuan, and McLeod 2012; Stasser and Birchmeier 2003).

Therefore, risk workshops can be framed as distributed cognition (Harten, Meyer, and Bellora-Bienengräber 2022). That is, the cognitive task is not performed by individuals in isolation but by a group as a whole, using the cognition and knowledge of all participants (Hauke, Lorscheid, and Meyer 2018). Our model implements transactive memory, as it allows participants to learn about and afterwards use the knowledge of other participants.

Merging this cognitive perspective with a discursive perspective, Harten, Meyer, and Bellora-Bienengräber (2022) identify several potential drivers of discussions' effectiveness in risk workshops from the literature. These drivers encompass (1) the effects of limits to information transfer within the group (i.e. knowledge of other participants' knowledge when integrating new information, instead of discarding previously held beliefs), (2) incomplete discussions (i.e. the rules by which it is decided to end the risk workshop instead of continuing the discussion), (3) group characteristics (i.e. information distribution, consideration of hierarchical relationships, and knowledge about each other's fields of expertise), and (4) the interaction patterns applied in the group (i.e. the order in which participants are allowed to talk).

Calculative cultures in risk management

Recently, corporate culture has received increasing attention as a key factor influencing risk management and its effectiveness. The term risk culture is used in different ways, among others, as 'the shared preferences towards risk and uncertainty' (Pan, Siegel, and Wang 2017, 2328) or, more specifically, as a subset of organizational culture, specifying how 'organisations think about, know, process and act upon risks and uncertainties' (Power 2020, 45). In this paper, we specifically focus on the latter and address the effects of different calculative cultures (Mikes 2009) that respectively represent different attitudes toward calculative practices.

Previous research identified two different calculative cultures in the context of two distinct approaches to enterprise risk management (ERM) in organizations.[7] Mikes (2009) studied ERM practices at two banking organizations and contrasted their ERM models as 'ERM by the numbers' and 'holistic ERM'. These different practices stem from different corporate governance pressures (Power 2007).

First, the ERM-by-the-numbers approach to managing risks aims at measuring the impact thereof on shareholder value. This leads to an emphasis of quantifiable risks, the impact of which on shareholder value is calculable. The overall risk portfolio is described as an aggregate and can be compared to the organization's risk appetite. In this manner, ERM contributes to the overall performance measurement. According to this perspective of risks, the calculated values for risks are the decisive output of the risk management process. The focus is on improving the quality of these calculated values by improving risk models. Obviously, this approach is ill-suited when dealing with hard-to-quantify risks. Mikes (2009) labels the calculative culture embedded in this type of ERM as 'quantitative enthusiasm'.

Second, the holistic ERM is not directly aimed at shareholder value but at achieving the organization's strategic objectives. So, ERM focuses on the identification of what it is that puts the achievement of those objectives at risk. While quantifiable risks are still relevant from this perspective, it also takes hard-to-quantify risks into account. ERM is not used to calculate overall risk exposure but to learn about the conditions under which the organization runs. The focus is less on precision and more on inclusiveness when understanding the overall risk environment. Mikes (2009) labels the calculative culture embedded in this ERM type as 'quantitative skepticism'.

The scope of the two ERM models obviously overlaps. The same risk can potentially be addressed by using either approach to ERM. Nevertheless, as both calculative cultures derive from different objectives, the reasoning about the risk will be different. We therefore discuss how different approaches to cognition represent these different calculative cultures in risk assessments.

Cognitive architectures of different calculative cultures

When simulating processes that, like risk workshops, involve human cognition, one must choose the most appropriate model of the cognition of the people involved. This requires the choice of a cognitive architecture. A cognitive architecture is a description of how information (including knowledge or beliefs) is stored in memory, how this memory is structured (i.e. the relationship between elements within the memory), and how the memory is processed (i.e. how it is utilized to learn or to reach conclusions) (Langley, Laird, and Rogers 2009; Thagard 2012). Two architectures used to investigate how humans make causal inferences are explanatory coherence and Bayesian networks (Thagard 2004).

Bayesian networks as a cognitive architecture

The ABM presented by Harten, Meyer, and Bellora-Bienengräber (2022) to model risk workshops uses Bayesian networks as the cognitive architecture of the agents. Bayesian networks model causal relationships between nodes that represent variables of interest (Pearl and Russell 2000). Each node is associated with a probability value, which represents the degree of belief regarding the corresponding variable's state (Thagard 2004). The connections between the nodes represent causal links. Each node's probability value is linked to connected nodes by dependent probabilities (e.g. 'if new competitors enter the market, how likely is it that they will target the same client segment'). The probability values can either be deduced by logical reasoning or by inference from observations of reality.

Given information about the true state of some of the nodes allows deducing the probability of all remaining nodes. The possibility of calculating specific probability values from causal relationships and limited information has made Bayesian networks a popular tool in risk assessment in particular and in artificial intelligence systems in general (Fenton and Neil 2019). Therefore, Bayesian networks are often used as a calculative tool to process information.

Explanatory coherence as a cognitive architecture

The theory of explanatory coherence was explicitly developed to explain why and how humans acquire certain beliefs. Among others, the theory has been implemented in the computational model ECHO (Thagard 1989). Like Bayesian networks, ECHO networks are built from interconnected nodes. The connections describe symmetric relationships between the nodes (e.g. 'our product has a high profit margin' is coherent with 'new competitors are attracted to our market'). Where Bayesian networks rely on specific probability values and allow precise calculations, ECHO is derived from basic principles[8] and is usually employed without adjusting concrete weights or otherwise providing numeric parameters to the network. An ECHO network can, therefore, usually be fully documented by a graph depicting the nodes and the relationships between them. Relationships between nodes are either explanatory (i.e. a 'high profit margin' is coherent with the 'market is attractive for competitors'), or contradictory (i.e. 'our main competitor failed to introduce a competing product' is incoherent with 'new competitors might emerge in our field'). Nodes represent either hypotheses or facts. The activation of each node has a numeric value. Like a system of interconnected springs, the network adjusts the activation of the nodes until it reaches a stable state, satisfying the constraints imposed by the explanatory and contradictory links (see Thagard 1989).

Calculative cultures: underlying cognitive architectures

Decision making in risk assessment involves the gathering of information and the identification of causal relationships, thereby deducing a risk assessment from available information (e.g. Fenton et al. 2020). While both Bayesian networks and explanatory coherence have been proposed as the appropriate cognitive architectures to model causal inference (Thagard 2004), both are, in principle, suited for this risk assessment task. We argue that the choice depends on the predominant calculative culture to be represented.

Bayesian networks require accurate probabilistic information and calculate a precise output (e.g. a specific value for a risk probability or impact). Such a model is akin to quantitative enthusiasm: every piece of information is quantifiable and the outcome can be used for further calculations.[9]

ECHO networks, by contrast, cannot incorporate precise values for probabilities. Instead, they rely on qualitative descriptions of relationships. They model how individuals make sense of information and account for the limited capability of individuals to make precise calculations or complex logical deductions. Thus, ECHO networks are well suited to model discussions in a quantitative skepticism setting. In line with this conceptual distinction, we compare results of the simulation of risk workshops generated by Harten, Meyer, and Bellora-Bienengräber (2022) using Bayesian networks to model a quantitative enthusiasm setting with results of the simulation of risk workshops using ECHO networks to model a quantitative skepticism setting.

Method

Overall design

Like Harten, Meyer, and Bellora-Bienengräber (2022), this study uses a simulation experiment approach which combines a model of the processes we want to investigate with an appropriate experimental design (Harrison et al. 2007). The ODD + D (Overview, Design Concepts and Details + Decision) protocol, containing a standardized description of the technical implementation of the simulation (Grimm et al. 2006, 2020; Müller et al. 2013), is reported in the Web Appendix (supplementary material) and on www.comses.net (the latter includes also the code of the simulation).

To better understand the impact of calculative culture on the outcome of risk workshops, we compare the results produced by Harten, Meyer, and Bellora-Bienengräber (2022) with those of an implementation using a different cognitive architecture. While we change the cognitive architecture used, we employ the same approach to model the interaction of the participants in the risk workshops. The interaction is modeled as an ABM with information exchange between agents (Harten, Meyer, and Bellora-Bienengräber 2022; Lorscheid and Meyer 2021; Wall and Leitner 2021).[10]

Model of the discussion process and risk assessment

The discussion process used for this study is identical to the one used by Harten, Meyer, and Bellora-Bienengräber (2022). For each simulation experiment, we conduct multiple simulation runs. Each run is a discussion of a single risk in a risk workshop. The simulation run consists of five stages (Figure 1).

In the ECHO networks used for this study, we adapt the structure of the Bayesian networks used by Harten, Meyer, and Bellora-Bienengräber (2022). Figure 2 compares the structure of the Bayesian network used by Harten, Meyer, and Bellora-Bienengräber (2022) with the ECHO networks used in this study. For each discussion, a new risk is generated. The network representing full knowledge of a risk contains 38 nodes, comprising 27 information nodes, nine issue nodes, and two nodes from which the overall risk assessment regarding the likelihood of the risk is derived (see Davies et al. 2010 for a similar conceptualization in the context of decision making in risk regulation). Due to differences in their backgrounds or priorities, individual participants start with diverse risk perceptions (Sjöberg 2000). Initially, they are only aware of the information and issues they are provided with before the start of the discussion. Information on the 36 information and issue nodes is exchanged during the discussion. If participants hear about a node that they have previously been unaware of, they include it in their mental model. All nodes have an activation between −1 and 1, representing their degree of belief in the underlying information or issue. Depending on their knowledge, participants can reach different risk assessments for the same risk.

Model of the discussion

Like in Harten, Meyer, and Bellora-Bienengräber (2022), nine participants exchange information in the risk workshop. These participants make their best effort to gain a correct understanding of the risk, i.e. they do not follow own and potentially hidden agendas. For example, there are no agency conflicts between participants at different hierarchical levels. The discussion consists of discussion rounds, each following a set sequence of participant activities (see stage 4 in Figure 1). The impact of the distinctive design characteristics of the risk workshop is analyzed in four simulation experiments.[11]

Simulation experiment 1: limits to information transfer

Receivers of information do not simply adopt the view of the sender of the information as to their own, but make sense of the information and integrate it into their respective mental models. Therefore, even if all information is shared during the risk workshop, participants will not necessarily have identical mental models at the end of the workshop. Simulating the information exchange in the risk workshop always allows us to investigate the mental state of all

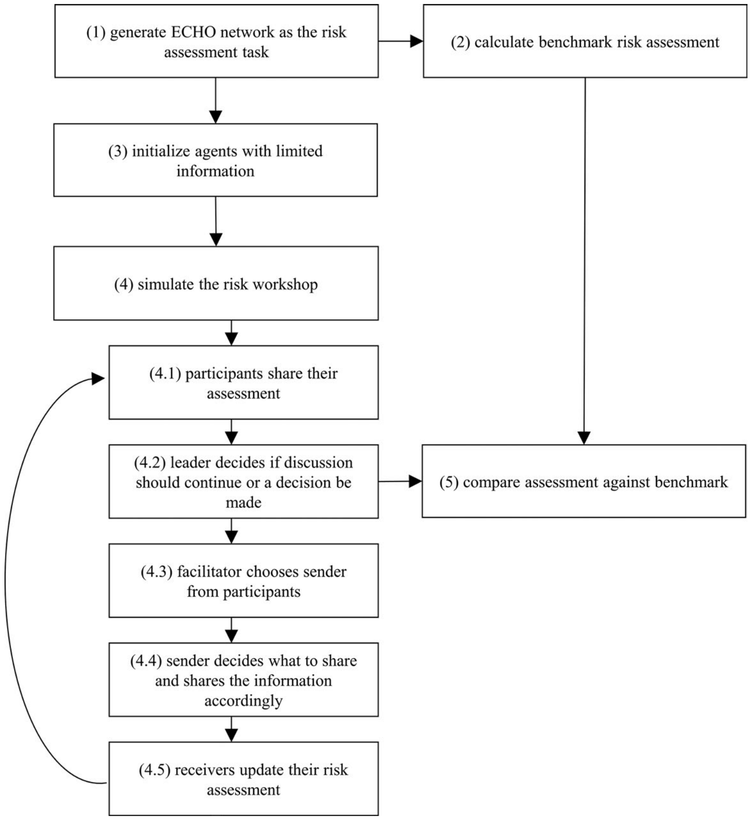

Figure 1. Stages of the simulation before, during, and after the risk workshop (adapted from Harten, Meyer, and Bellora-Bienengräber 2022).

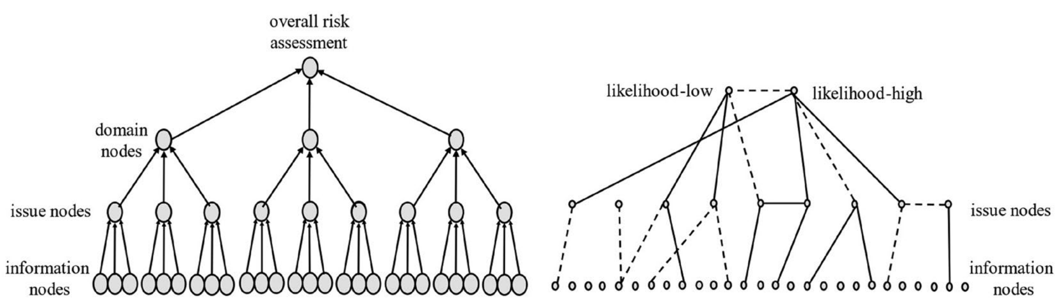

Figure 2. The structure of the Bayesian network used by Harten, Meyer, and Bellora-Bienengräber (2022) (left) and the stylized ECHO network used to model quantitative skepticism (right). *Note:* For the ECHO network, full lines indicate explanatory relationships and dotted lines contradictory relationships. For visual clarity, only a selection of the existing relationships is depicted. Relationships can exist between information and issue nodes, among the issue nodes, and between issue nodes and the likelihood nodes used for deriving the risk assessment.

participants and thereby investigate the evolvement of the risk assessment of the individuals and, accordingly, of the group.

Simulation experiment 2: incomplete discussions

Every discussion must stop after the elapse of a certain amount of time. In the simulated risk workshop, the decision to end the discussion and decide on a risk assessment is made by the leader (i.e. a designated participant). The leader uses heuristics to determine whether the discussion should end. Harten, Meyer, and Bellora-Bienengräber (2022) investigate two heuristics: either the leader requires a consensus of all participants to end the discussion, or the discussion ends after it has stagnated for some time. Here, stagnation is defined as no change of the average (numerical) risk assessment over several rounds.

While Harten, Meyer, and Bellora-Bienengräber (2022) use stagnation in the discussion as an indicator to terminate it, this is not suitably applicable to the quantitative skepticism model. In the quantitative enthusiasm model the assessments tend to converge on the final risk assessment, and a slow rate of change denotes reaching the final stage of the discussion. The same does not apply to the quantitative skepticism model. Here assessment changes happen suddenly, even after long periods of unchanged risk assessments. We therefore also simulate another heuristic, i.e. the continuation of the discussion until each piece of information has been mentioned at least once.

Simulation experiment 3: specific group characteristics

We model the effect of the same three group characteristics analyzed by Harten, Meyer, and Bellora-Bienengräber (2022), thus accounting for the following group characteristics:

- Differences in the distribution of information. Information is distributed uniformly among the participants before the start of the discussion, or some participants are initially provided with more information than others.
- Differences in hierarchy. In the baseline scenario, participants disregard information shared by someone higher up or lower down in the hierarchy. Alternatively, they can weigh the provided information to reflect the higher belief attached to the information by someone higher up in the hierarchy.
- Information about each other's field of expertise (transactive memory). When participants have a transactive memory, they can give greater weight to information provided by experts. Otherwise, they can dismiss the expertise when processing the provided information.

Simulation experiment 4: specific interaction patterns

Like Harten, Meyer, and Bellora-Bienengräber (2022), we simulated five patterns according to which the risk workshop facilitator arranges the order of speakers during the workshop. The baseline scenario assumes that the next speaker is chosen at random. In the remaining patterns the facilitator selects participants who are concerned about the risk (i.e. participants who assess the risk higher than other participants); participants whose assessment differs the most from the average assessment of the other participants (i.e. dissenters); participants whose assessment is often close to the average assessment of the other participants (i.e. when homogeneity prevails); and, lastly, the facilitator prioritizes participants based on their higher hierarchical position.

Benchmarking the risk workshop

As a baseline for evaluating the effectiveness of a risk assessment workshop, it is necessary to define a benchmark risk assessment. Harten, Meyer, and Bellora-Bienengräber (2022) propose a simulated risk workshop under ideal conditions as the benchmark risk assessment. This choice rests on the assumption that performing a discussion using settings of an ideal speech situation—where experts share only correct information, and all other participants integrate the correct information into their mental model—leads each participant to the correct risk assessment. Indeed, for their quantitative enthusiasm model, Harten, Meyer, and Bellora-Bienengräber (2022) show this effect in their benchmark simulation.

In the quantitative skepticism model, we find that the order in which participants learn new information strongly influences the risk assessment they reach, both individually and collectively. Furthermore, depending on their current mental model (i.e. their individual ECHO model when they learn new information), the participants react differently to new information. As there is no clear-cut, 'ideal' order of receiving new information, there is also no mechanism that always grants all participants the collective ability to reach an identical assessment that could serve as a benchmark. Consciously deviating from Harten, Meyer, and Bellora-Bienengräber (2022), we therefore use the assessment of a hypothetical agent that has all information (i.e. the complete ECHO network representing the risk) from the outset, without any stepwise learning through the risk workshop discussion, as a benchmark risk assessment. This state is analogous to the consensus assessment reached under ideal conditions in the quantitative enthusiasm model of Harten, Meyer, and Bellora-Bienengräber (2022) as, for our model, having all information at the outset is conceptually equivalent to the result of learning all information under ideal discourse conditions.

Results and discussion

Individual dynamics of the discussion

Figure 3 depicts two examples of the evolvement of the individual risk assessment of the nine participants during a simulated risk workshop. Although, in both examples, the participants simultaneously receive the same information during the discussion, they reach different conclusions. These assessments are based on the previous state of their respective mental models and can differ accordingly (e.g. information might shift the assessment of one participant but be absorbed by another participant's mental model that is consistent with the new information). During certain risk workshops like the one depicted on the left-hand side of Figure 3, all participants will reach the same overall risk assessment. In other workshops, like the one depicted on the right-hand side of Figure 3, participants split into distinct groups, even after sharing all information.[12]

Given this strong dependency of the risk assessment on the risk workshop participant's previous mental model (and thus on the timing of receiving the information), we cannot expect all participants to reach a consensual risk assessment under these conditions. Thus, to define a benchmark risk assessment, we use the risk assessment made by an agent having all information from the outset (see the *Method* section).

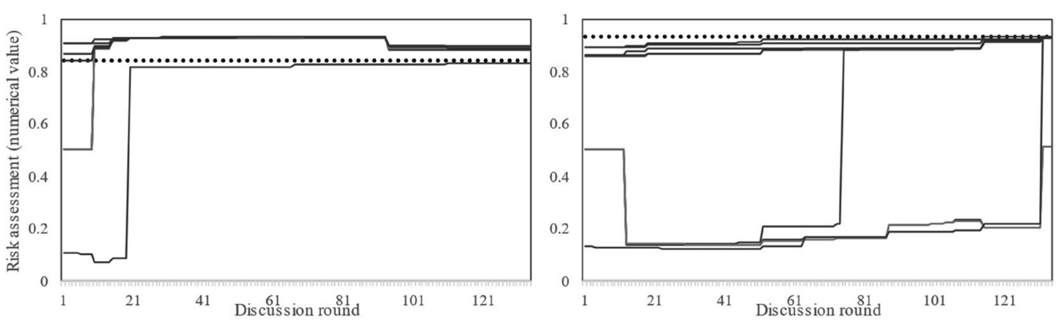

Figure 3. Two examples of individual risk assessments of all nine participants during the risk workshop. *Note:* The dotted line marks the benchmark risk assessment. The left-hand panel shows a risk workshop where participants reach an assessment close to the benchmark, early in the discussion. The right-hand panel shows a risk workshop where participants change their assessment late in the discussion.

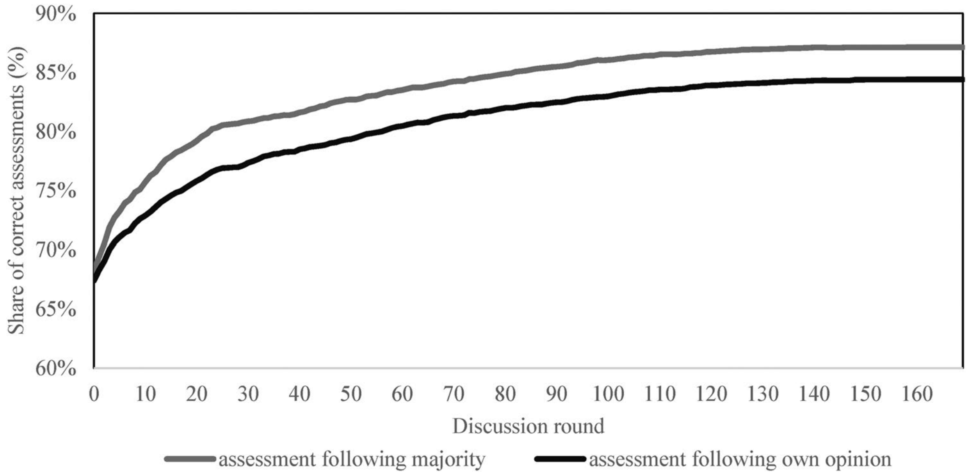

Figure 4. Development of the correctness of the risk assessment over discussion rounds. *Note:* The figure depicts the share of risk workshops that would reach a correct risk assessment if ended after the current discussion round, following either the majority opinion or the leader's own opinion.

Collective dynamics of the discussion

Simulation experiment 1: limits to information transfer

As shown in the previous section, under a quantitative skepticism culture, changes in individual risk workshops occur as dramatic shifts in single opinions (see the upward and downward movements in the risk assessment in Figure 3). However, in the aggregate over many risk workshops, we find that correctness increases continuously. This is clear in Figure 4, which depicts the share of risk workshops that would achieve a correct risk assessment after a certain number of discussion rounds when the leaders follow either their own opinion or the assessment of the majority of the participants. Accordingly, we find that both the leader's own assessment and the majority assessment on average continuously improve the correctness of all simulated risk workshops until reaching an upper limit.[13] In comparison, under the condition of quantitative enthusiasm (Harten, Meyer, and Bellora-Bienengräber 2022), individual discussions evolve more gradually. As a result, the aggregated correctness over many risk workshops nevertheless increases faster at the beginning of the risk workshop, with diminishing correctness returns over time. Hence, when a quantitative skepticism culture prevails, it is reasonable—from the viewpoint of improving the correctness of the resulting risk assessment—to continue the discussion until all information has been shared (i.e. after 118 ± 30 discussion rounds; untabulated). Still, even after such a lengthy discussion, a correct risk assessment is not always achieved. However, when a quantitative enthusiasm culture dominates and the participants' risk assessment evolves more gradually, a few stable rounds (after which all information has not yet been shared) may be sufficient to correctly classify a large number of risks.

Simulation experiment 2: incomplete discussions

In our model, we find that 60% of discussions (untabulated) reach a consensus and that this happens, on average, after 33.4 discussion rounds (Table 1). In these cases, the consensus is correct for 98.1% of all risks. If the discussion is ended based on its stagnation, we find that following the group majority has a slight but consistent advantage across all risks, compared to merely following the leader's individual assessment, independent from the required length of stagnation (either one, five, or ten discussion rounds without a change in the average risk assessment).

Table 1. Effects of incomplete discussions on risk assessment effectiveness.

Decision-making rule	Proportion of correct assessments: All risks (%)	High risks (%)	Low risks (%)	Avg. number of discussion rounds
Stop at first group consensus	98.1%	98.0%	98.2%	33.4
One stable round				
Leader follows own opinion	68.8%	87.1%	49.7%	2.1
Leader follows majority	70.5%	95.8%	43.9%	
Five stable rounds				
Leader follows own opinion	71.8%	84.4%	58.6%	7.6
Leader follows majority	74.7%	91.6%	57.1%	
Ten stable rounds				
Leader follows own opinion	73.2%	83.3%	62.7%	15.5
Leader follows majority	78.3%	88.5%	67.6%	
Full discussion				
Leader follows own opinion	83.0%	85.3%	80.6%	117.8
Leader follows majority	87.0%	86.3%	87.7%	

Note: The table shows the percentage of risks that are correctly assessed and the average number of rounds before the decision is made, depending on the mechanism used to end the discussion. A discussion is said to have *n* stable rounds if the average (numerical) risk assessment does not change more than 2% over *n* consecutive rounds.

If the discussion is only terminated when every piece of information has been discussed at least once, 87% of all risks are correctly assessed, though on average only after 118 discussion rounds.[14] When comparing the results after one stable round with the results after ten stable rounds, it again becomes apparent that, in a quantitative skepticism culture, a temporarily stagnant discussion is not a good indicator that the risk workshop has reached its best achievable result (see the results above). Due to the comparatively dramatic shifts in opinions particular to this calculative culture, it is necessary to continue the discussion through these phases of (apparent) stagnation to reach the best possible results.

Harten, Meyer, and Bellora-Bienengräber (2022) identify a pattern where initially learning about the risks structure increases uncertainty. This mechanism is not present in the quantitative skepticism model, as learning about new information does not introduce uncertainty.[15] New information will be processed and at once converted into a state of coherence with previously available information. What is important in this model is the integration of the information into the mental model, and not so much the extent to which the information is believable. For example, the idea that a new competitor might enter the market has a greater impact than the actual probability thereof. However, like Harten, Meyer, and Bellora-Bienengräber (2022), we find that participants can also be 'right for the wrong reason', as they also have a slight tendency to initially overestimate risks. Participants with too little information to inform their risk assessment will default to a 'high' assessment. Subsequently, they will adjust their risk assessment downwards as they learn more, thereby reducing the rate of correctly assessed high risks and increasing the rate of correctly assessed low risks. Herein lies a tradeoff between not identifying high risks and overestimating too many low risks.

Simulation experiment 3: group characteristics

Concerning group characteristics, we observe the highest effectiveness of risk assessments when information is equally distributed among the participants (Table 2).[16] This is in line with the results obtained by Harten, Meyer, and Bellora-Bienengräber (2022) in a quantitative enthusiasm setting. We do not find a significant impact (untabulated) of the presence of transactional memory or of taking hierarchical differences into account. In the quantitative skepticism model, compared to the quantitative enthusiasm model, it is less important how strong the first belief in any particular piece of information is (which could be impacted by

Table 2. Effects of group characteristics on risk assessment effectiveness.

Differences in the distribution of knowledge	Receivers consider hierarchical differences	Receivers have no transactive memory	Proportion of correct assessments: All risks (%)	High risks (%)	Low risks (%)	Avg. number of discussion rounds
+	+	+	83.7%	85.4%	81.9%	136.9
+	+	−	80.2%	80.3%	80.1%	136.1
+	−	+	81.3%	80.7%	81.9%	136.3
+	−	−	81.6%	80.3%	83.0%	135.7
−	+	+	85.7%	85.2%	86.2%	117.3
−	+	−	87.4%	87.2%	87.6%	118.7
−	−	+	86.2%	83.3%	89.2%	118.2
−	−	−	87.4%	87.5%	87.3%	118.1

Note: The table shows the percentage of correctly assessed risks after a full discussion when the leader follows the majority vote and the average number of rounds before the decision is made, depending on group characteristics that might influence the risk workshop's effectiveness. A '+' indicates the presence of the corresponding deviation from an ideal discussion situation. Differences in the distribution of information are implemented by giving some participants a higher probability to receive information during the initialization. If receivers consider hierarchical differences, they weigh input according to the sender's hierarchical position compared to their own. With no transactive memory, participants do not distinguish between senders who are experts on the information and those who are not.

Table 3. Effects of interaction pattern on risk assessment effectiveness.

Who talks next during the discussion?	Proportion of correct assessments: All risks (%)	High risks (%)	Low risks (%)	Avg. number of discussion rounds
Random choice of participants	87.4%	87.5%	87.3%	118.1
Priority to concerned participants	81.6%	95.1%	69.0%	155.7
Priority to participants with dissenting opinions	83.8%	82.1%	85.5%	128.8
Priority to participants with higher hierarchical position	87.8%	87.5%	88.2%	127.1
Priority to participants close to group opinion	86.7%	84.9%	88.5%	737.9

Note: Percentage of risks that are correctly assessed after a full discussion when the leader follows the majority vote and the average number of rounds before the decision is made, depending on the interaction pattern. Concerned participants are those who assess the risk as particularly high. Dissenting participants and those close to the group opinion are determined by measuring the distance between their risk assessment and the average risk assessment of the group.

the hierarchical position or expertise of the sender), as the belief will immediately be adjusted to a coherent state regarding other information already available to the participant. For example, as soon as the issue that 'a competitor might enter our market' is introduced to the mental model, it might affect the participant's risk assessment, even if the activation—i.e. the degree of belief—is low because, for instance, the information stems from a peer instead of from a superior.

Simulation experiment 4: interaction patterns

Only the interaction pattern favoring concerned participants differs remarkably from the baseline interaction scenario following a random order of participants (Table 3).[17] Favoring concerned participants will improve the assessment of high risks but will decrease the rate of correctly identified low risks. Thus, this interaction pattern introduces a bias toward assessing risks as high. Our results differ from Harten, Meyer, and Bellora-Bienengräber (2022) regarding the assessment of low risks with prioritized, concerned participants, as they find an improvement in the assessment of all risks, as opposed to the better assessment of high risks only. Therefore, prioritizing concerned participants comes at the cost of correctly assessing low risks in a quantitative skepticism model. Given such a rule, this results from the focus on high-risk assessments during the early discussion rounds of the risk workshops.

For rules prioritizing hierarchy and homogeneity, Harten, Meyer, and Bellora-Bienengräber (2022) also identify significant changes compared to the baseline scenario. These are less pronounced or even absent in the quantitative skepticism model, as it is less important who first introduces a new piece of information as long as it is introduced. Participants will at once assess the information's consistency with their previous mental model. It is also noteworthy that prioritizing participants close to the group opinion will dramatically increase the number of discussion rounds needed to introduce all information, while in Harten, Meyer, and Bellora-Bienengräber (2022), the discussions are ended earlier due to the termination criteria based on sensible stagnation in this quantitative enthusiasm culture.

Summary and comparison of calculative cultures

By repeating the simulation experiments performed by Harten, Meyer, and Bellora-Bienengräber (2022) and using a model representing a different calculative culture, we find that the drivers of the effectiveness of risk assessments are partially sensitive to the dominant calculative culture. Table 4 summarizes the results of both simulation studies and thereby highlights similarities and differences. We learn from simulation experiment 1 that, notably for both calculative cultures, the risk assessment improves during the discussion. Still, path-dependencies are

Table 4. Comparison of results for the simulation experiments with quantitative enthusiasm and quantitative skepticism.

	Calculative culture	
	Quantitative enthusiasm	Quantitative skepticism
Process	Overall improvement of risk assessments over time	Improvement of risk assessments over time
	A tradeoff between the identification of high and low risks	A tradeoffs between the identification of high and low risks
	Gradual changes in individual and collective assessments	Sudden and unpredictable shifts in individual and collective assessments
Design		
Incomplete discussions	Stagnation in assessments as an indicator to end the discussion	Changes in assessment happen suddenly. Continue discussions even after lengthy periods of stagnation until all information have been shared
		Following the group majority always has a slight but significant advantage across all risks
Group characteristics	Differences in the distribution of information have negative effects	Differences in the distribution of information have significant negative effects
	The absence of transactive memory has negative effects	The absence of transactive memory has no significant negative effects
	Hierarchical differences have negative effects	Hierarchical differences have no significant effects
Interaction patterns	Prioritizing concerned participants improves the assessment of all risks	Prioritizing concerned participants improves the assessment of high risks but lowers the rate of correctly identified low risks
	Prioritizing dissenting participants improves the assessment of high risks but lowers the rate of correctly identified low risks	Prioritizing dissenting participants lowers the rate of correctly assessed risks
	Significant changes compared to the baseline scenario for prioritizing dissent, hierarchy, and homogeneity	No observed effects for prioritizing hierarchy and homogeneity

Note: The results for quantitative enthusiasm are based on Harten, Meyer, and Bellora-Bienengräber (2022), while the results for quantitative skepticism are based on this study. Our simulation experiments were designed to allow a comparison.

characteristic of discussion processes given quantitative skepticism, i.e. it does not only matter *what* information is exchanged but also *when*. Moreover, there is a potential tradeoff between correctly identifying high and low risks in both calculative cultures, as the rate of correct assessments of certain risks decreases during the workshop.

Simulation experiment 2 shows that, with quantitative skepticism, it is more difficult to find the correct time to end the discussion, as the rate of correct risk assessments continuously improves until all information has been shared; stagnation is not a good indicator that the discussion is over. Simulation experiment 3 showcases that, while both calculative cultures show better results with an equal distribution of information within the group, quantitative enthusiasm is not negatively affected by a lack of transactive memory or the presence of hierarchical differences. It only matters that all information is made available. The credibility or hierarchical position of the participant introducing the information is less important than the compatibility of the information with the mental model of the receivers.

Regarding the interaction patterns, simulation experiment 4 provides evidence that prioritizing concerned participants will improve the assessment of high risks for both calculative cultures. However, when quantitative skepticism dominates, there is a tradeoff with correctly assessing low risks. Otherwise, settings with quantitative skepticism are less impacted by interaction patterns than settings with quantitative enthusiasm.

Conclusion

Calculative cultures affect how risk-related information is processed by individuals and translated into risk assessments. This makes it a critical boundary condition for the effectiveness of the design and implementation choices in risk workshops—a common risk assessment technique in organizations. This paper explored drivers of the effectiveness of risk assessments in risk workshops, given a calculative culture of quantitative skepticism, and compared them with the findings of previous research that implicitly assumed a calculative culture of quantitative enthusiasm. Using ABM, we modeled individuals' information processing and judgment formation and represented this calculative culture using ECHO models capturing agents' cognition. This allowed us to extend previous simulation experiments of ABM-rooted risk workshops by consciously incorporating the role of the predominant calculative culture in the modeling.

Our results make three contributions to research and practice. First, given the calculative nature of quantitative skepticism, some distinct effects must be considered when designing and implementing risk workshops. Notwithstanding Mikes (2009) early recognition of the importance of distinguishing between different calculative cultures, our study is, to the best of our knowledge, the first to show the considerable impact that differences in calculative cultures have on the outcome of risk assessments. We show that the predominant calculative culture of an organization is a critical boundary condition when considering the process of risk workshops over time.

Compared to Harten, Meyer, and Bellora-Bienengräber (2022) findings in a quantitative skepticism culture, the improvement of the correctness of the risk assessment is anything but a gradual process; we rather document sudden shifts in individual and collective assessments. This result questions whether, given this boundary condition, a stagnating discourse is a good indicator to end a discussion. Instead, risk workshop facilitators should create a setting in which everyone is encouraged to share all their information as soon as possible.

We also follow the call of Katzenbach and Smith (2015) to specify rules of interaction. Unlike Harten, Meyer, and Bellora-Bienengräber (2022), we do not observe positive effects when prioritizing dissent or negative effects of hierarchy or homogeneity. Additionally, we are the first to document a possible bias when applying the common rule of thumb to prioritize concerned participants during the risk workshop, as supported by Harten, Meyer, and Bellora-Bienengräber (2022). We find that favoring concerned participants improves the assessment of high risks but that it lowers the rate of correctly identified low risks. This result indicates the possible existence of a focusing illusion or an anchoring bias (Kahneman 2011), also for risk workshops. In contrast to Moreland and Myaskovsky (2000) and what is found in Harten, Meyer, and Bellora-Bienengräber (2022), we observe no positive effect of a group member's familiarity with the other members' expertise on group performance. Facilitators should be aware of the predominant calculative culture in the organization when deciding whether they need to limit the involvement of dissenters and superiors. Overall, risk workshop facilitators should gain expertise in recognizing the predominant calculative culture in an organization.

Second, we show that some recommendations by Harten, Meyer, and Bellora-Bienengräber (2022) are robust and therefore resistant to different calculative cultures. Previous research, assuming quantitative enthusiasm, documented a tradeoff between the correct assessment of high and low risks (Harten, Meyer, and Bellora-Bienengräber 2022). This tradeoff is also present in a predominant quantitative skepticism culture. Facilitators need to be aware that any decision on the design and implementation of risk workshops must take the organization's risk appetite into account. Increasing the correctness of detecting high risks will come at the cost of overestimating low risks, independent of the organization's positioning along the calculative culture continuum. Moreover, we also find that an equal distribution of information within the group leads to the most effective risk workshops in both calculative cultures.

Third, we contribute conceptually and methodologically to the risk literature by introducing a novel approach that allows a model-based investigation of two distinctive ways risk information is processed and used for risk assessments. So far, the literature has mainly used case studies to identify related models (Mikes 2009). In addition, literature used surveys to quantify the antecedents and consequences of related types of information usage, such as the diagnostic vs. interactive use of information (Simons 1990) or the use of accounting practices as a computation tool instead of a learning tool (Burchell et al. 1980). While the fine-grained study of the actual usage of information and related cognitive processes is particularly challenging in case studies and even more so in surveys, it can be made amenable for a detailed analysis through ABM. Our ABM, made replicable by our in-depth description in the ODD + D, enables future researchers to model the complex interplay between calculative culture, individual cognitive processes, and the related group-level outcomes. Moreover, the scant prior ABM modeling approaches of risk assessments (Harten, Meyer, and Bellora-Bienengräber 2022) do implement but not conceptualize the modeling of a specific calculative culture. Therefore, to the best of our knowledge, this is the first study on risk assessment that conceptualizes and investigates the role of calculative cultures for individual cognitive processes in combination with other organizational variables.

Like other studies, this paper has limitations that future research could address. First, culture is a complex and rich phenomenon. Although we believe that our model represents key facets, a formal representation of calculative cultures via different cognitive architectures cannot capture all nuances of reality. Future research should draw on the rich qualitative work of Mikes (2009) and others to capture facets not included in our study, like the maturity of the risk function. Second, to allow a comparison of the two calculative cultures identified by Mikes (2009), we followed the experimental design of Harten, Meyer, and Bellora-Bienengräber (2022) by focusing on the assessment of a single risk and by making a binary distinction between high and low risks. In addition, this also includes their choice of manipulated variables. Future research should search for additional factors that explain differences in the effectiveness of risk assessments in risk workshops. For example, as suggested by Bromiley et al. (2015), this could include participants' motivation, hidden agendas, or changes in the groups' ability to transfer the information as the discussion progresses. As ABM allows the incorporation of different personality types (Davies et al. 2010), researchers should also investigate the effects of different portions of extroverted vs. introverted workshop participants and how this interacts with calculative culture and the other risk workshop design characteristics considered in this paper. Third, our model is limited to the context of risk workshops taking place with participants participating synchronously in the discussion. Thus, it does not include risk assessments that happen with dispersed participants interacting asynchronously with each other (e.g. via e-mail). While it is likely that many effects identified in this paper would manifest also when considering the latter type of risk assessments, future research is called to broaden the applicability of our model by incorporating for example the effect of time delays in the communication or of selective attention to certain pieces of information. Fourth, the external validity of our results could be strengthened by future research through empirical tests in the form of field and case studies. It might, for example, be interesting to study a potential clash of different calculative cultures in the same risk workshop or organization. Finally, the use of the approach developed in this paper to investigate the effects of different calculative practices in other contexts beyond risk workshops and even beyond risk management, appears to be promising.

Notes

1. These calculative cultures are embedded in broader enterprise risk management (ERM) practices. While quantitative enthusiasm is a cornerstone of an 'ERM by the numbers', quantitative skepticism is a descriptive component of a 'holistic ERM' (Mikes 2009, 35).
2. We consider these two calculative cultures as two extreme points of a continuum, similar to Mikes' (2009) suggestion that the related ERM models might be 'different stages in the evolution of risk management' (Mikes 2009, 37).
3. 'Risk' means uncertainty about how certain events, when they occur, may affect the organization. These events can have positive and negative outcomes (COSO 2017). In this paper, similar to Harten, Meyer, and Bellora-Bienengräber (2022), we restrict ourselves to those risks that may result in negative outcomes (COSO 2017). However, our modeling applies to both threats and opportunities.
4. In the following, 'information' refers to the participant's organized data in the context of the risk assessment task, while 'knowledge' refers to cognitively processed and aggregated information that enables participants to reach an understanding of the assessed risk.
5. Our modeling can be also used in future research to incorporate the impact dimension of risk assessments.
6. Like Harten, Meyer, and Bellora-Bienengräber (2022), we also investigate the design and implementation of workshops from the point of view where the worst credible impact of a particular risk is clear. So, the assessment focuses on the likelihood of the risk's worst credible impact. Subsequently, to ensure clarity, 'high risks' and 'low risks' refer to 'high likelihood risks' and 'low likelihood risks', respectively, and 'risk assessment' refers to the 'assessment of the likelihood of a risk'.
7. See also Arena, Arnaboldi, and Azzone (2011) for empirical data on the interplay between ERM implementation and risk culture.
8. For a definition of the seven basic principles of ECHO (i.e. symmetry, explanation, analogy, data priority, contradiction, competition, and acceptance) see Thagard (1989).
9. See Neil et al. (2021) for an example of how Bayesian networks can assist communication in a quantitative enthusiasm context.
10. See Davies et al. (2010) for a discussion of possible applications of ABM to model risk regulation.
11. An overview of the simulation experiments is provided in Table A.1.
12. This pattern is frequently observed in models of opinion dynamics, e.g. Hegselmann and Krause (2002).
13. Before the start of the risk workshop, participants can only rely on the information they are initially provided with. Without any information, a correct assessment—as a purely random choice between two assessments, i.e. 'high' and 'low'—is expected for 50% of the risks. However, the pieces of information provided before the start of the discussion (i.e. in discussion round 0) allow participants to make a correct risk assessment for about 69% of all risks.
14. The number of discussion rounds is very high compared to the numbers reported by Harten, Meyer, and Bellora-Bienengräber (2022), because we focus our investigation on risk workshops that only finish after all information has been shared and not, as Harten, Meyer, and Bellora-Bienengräber (2022), on risk workshops that finish after ten stable rounds.
15. In the quantitative enthusiasm model, learning the latest information occurs at two levels. At the first level, participants learn about the existence of a certain node, but not yet about its related probability (i.e. they learn what they do not know). At the second level, they assign a non-zero probability to the possibility that the node is in a 'high risk' state and their uncertainty increases. In later steps, participants attach a probability to the node. The learning at two levels does not occur in a quantitative skepticism model, where learning about the latest information does not introduce uncertainty.
16. Table A.2 presents simulation results of experiment 3 after 10 stable discussion rounds.
17. Table A.3 presents simulation results of experiment 4 after 10 stable discussion rounds.

Acknowledgments

We thank the participants of the 9th Annual Conference on Risk Governance (2021), the Social Simulation Conference (2019), and the Hamburg Management Accounting Research Seminar for their valuable comments and suggestions. We also thank the anonymous reviewers and the editor for their support and valuable comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Lucia Bellora-Bienengräber http://orcid.org/0000-0003-2013-0438
Matthias Meyer http://orcid.org/0000-0003-2980-4670

References

Arena, M., M. Arnaboldi, and G. Azzone. 2011. "Is Enterprise Risk Management Real?" *Journal of Risk Research, Routledge* 14 (7): 779–797. doi:10.1080/13669877.2011.571775.

van Asselt, M. B. A., and O. Renn. 2011. "Risk Governance." *Journal of Risk Research* 14 (4): 431–449. doi:10.1080/13669877.2011.553730.

Boholm, Å., and H. Corvellec. 2016. "The Role of Valuation Practices for Risk Identification." In *Riskwork*, edited by M. Power, 110–129. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198753223.003.0006.

Bromiley, P., M. McShane, A. Nair, and E. Rustambekov. 2015. "Enterprise Risk Management: review, Critique, and Research Directions." *Long Range Planning* 48 (4): 265–276. doi:10.1016/j.lrp.2014.07.005.

Burchell, S., C. Clubb, A. Hopwood, J. Hughes, and J. Nahapiet. 1980. "The Roles of Accounting in Organizations and Society." *Accounting, Organizations and Society* 5 (1): 5–27. doi:10.1016/0361-3682(80)90017-3.

COSO. 2017. *Enterprise Risk Management: Aligning Risk with Strategy and Performance*. New York, NY: COSO.

Davies, G. J., G. Kendall, E. Soane, J. Li, F. Charnley, and S. J. T. Pollard. 2010. "Regulators as 'Agents': power and Personality in Risk Regulation and a Role for Agent-Based Simulation." *Journal of Risk Research, Routledge* 13 (8): 961–982. doi:10.1080/13669877.2010.486075.

Fenton, N. E., and M. Neil. 2019. *Risk Assessment and Decision Analysis with Bayesian Networks*. 2nd ed. Boca Raton: CRC Press, Taylor & Francis Group.

Fenton, N. E., M. Neil, M. Osman, and S. McLachlan. 2020. "COVID-19 Infection and Death Rates: The Need to Incorporate Causal Explanations for the Data and Avoid Bias in Testing." *Journal of Risk Research* 23 (7–8): 862–864. doi:10.1080/13669877.2020.1756381.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, et al. 2006. "A Standard Protocol for Describing Individual-Based and Agent-Based Models." *Ecological Modelling* 198 (1–2): 115–126. doi:10.1016/j.ecolmodel.2006.04.023.

Grimm, V., S. F. Railsback, C. E. Vincenot, U. Berger, C. Gallagher, D. L. DeAngelis, B. Edmonds, et al. 2020. "The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism." *Journal of Artificial Societies and Social Simulation* 23 (2): 7. doi:10.18564/jasss.4259.

Harrison, J. R., Z. Lin, G. R. Carroll, and K. M. Carley. 2007. "Simulation Modeling in Organizational and Management Research." *Academy of Management Review* 32 (4): 1229–1245. doi:10.5465/amr.2007.26586485.

Harten, C., M. Meyer, and L. Bellora-Bienengräber. 2022. "Talking about the Likelihood of Risks: An Agent-Based Simulation of Discussion Processes in Risk Workshops." *Journal of Accounting & Organizational Change* 18 (1): 153–173. doi:10.1108/JAOC-11-2020-0197.

Hauke, J., I. Lorscheid, and M. Meyer. 2018. "Individuals and Their Interactions in Demand Planning Processes: An Agent-Based, Computational Testbed." *International Journal of Production Research* 56 (13): 4644–4658. doi:10.1080/00207543.2017.1377356.

Hegselmann, R., and U. Krause. 2002. "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation." *Journal of Artificial Societies and Social Simulation* 5 (3).

Kahneman, D. 2011. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Katzenbach, J. R., and D. K. Smith. 2015. *Wisdom of Teams: Creating the High-Performance Organization*. Boston, Massachusetts: Harvard Business Review Press.

Langley, P., J. E. Laird, and S. Rogers. 2009. “Cognitive Architectures: Research Issues and Challenges.” *Cognitive Systems Research* 10 (2): 141–160. doi:10.1016/j.cogsys.2006.07.004.

LiCalzi, M., and O. Surucu. 2012. “The Power of Diversity over Large Solution Spaces.” *Management Science* 58 (7): 1408–1421. doi:10.1287/mnsc.1110.1495.

Lorscheid, I., B.-O. Heine, and M. Meyer. 2012. “Opening the ‘Black Box’ of Simulations: increased Transparency and Effective Communication through the Systematic Design of Experiments.” *Computational and Mathematical Organization Theory* 18 (1): 22–62. doi:10.1007/s10588-011-9097-3.

Lorscheid, I., and M. Meyer. 2021. “Toward a Better Understanding of Team Decision Processes: combining Laboratory Experiments with Agent-Based Modeling.” *Journal of Business Economics* 91 (9): 1431–1467. doi:10.1007/s11573-021-01052-x.

Lu, L., Y. C. Yuan, and P. L. McLeod. 2012. “Twenty-Five Years of Hidden Profiles in Group Decision Making: A Meta-Analysis.” *Personality and Social Psychology Review* 16 (1): 54–75. doi:10.1177/1088868311417243.

McNamara, G., and P. Bromiley. 1997. “Decision Making in an Organizational Setting: Cognitive and Organizational Influences on Risk Assessment in Commercial Lending.” *Academy of Management Journal* 40 (5): 1063–1088. doi:10.2307/256927.

Mikes, A. 2009. “Risk Management and Calculative Cultures.” *Management Accounting Research* 20 (1): 18–40. doi:10.1016/j.mar.2008.10.005.

Moreland, R. L., and L. Myaskovsky. 2000. “Exploring the Performance Benefits of Group Training: Transactive Memory or Improved Communication?” *Organizational Behavior and Human Decision Processes* 82 (1): 117–133. doi:10.1006/obhd.2000.2891.

Müller, B., F. Bohn, G. Dreßler, J. Groeneveld, C. Klassert, R. Martin, M. Schlüter, et al. 2013. “Describing Human Decisions in Agent-Based models – ODD + D, an Extension of the ODD Protocol.” *Environmental Modelling and Software* 48: 37–48. doi:10.1016/j.envsoft.2013.06.003.

Neef, D. 2005. “Managing Corporate Risk through Better Knowledge Management.” *The Learning Organization* 12 (2): 112–124. doi:10.1108/09696470510583502.

Neil, M., N. Fenton, M. Osman, and D. Lagnado. 2021. “Causality, the Critical but Often Ignored Component Guiding us Through a World of Uncertainties in Risk Assessment.” *Journal of Risk Research* 24 (5): 617–621. doi:10.1080/13669877.2019.1604564.

Pan, Y., S. Siegel, and T. Y. Wang. 2017. “Corporate Risk Culture.” *Journal of Financial and Quantitative Analysis* 52 (6): 2327–2367. doi:10.1017/S0022109017000771.

Pearl, J., and S. Russell. 2000. “Bayesian Networks.” In *Handbook of Brain Theory and Neural Networks, edited by M.A. Arbib, 157-160*. Cambridge, MA: MIT Press.

Power, M. 2007. *Organized Uncertainty: Designing a World of Risk Management*. Oxford; New York: Oxford University Press.

Power, M. 2020. “Risk Culture and Information Culture: Why an ‘Appetite for Knowledge’ Matters.” In *Beyond Bad Apples: Risk Culture in Business*, edited by D. Ralph, K. Alexander, M. Tuveson, 42–72. Cambridge: Cambridge University Press. doi:10.1017/9781316996959.003.

Quail, R. 2011. “How to Plan and Run a Risk Management Workshop.” In *Enterprise Risk Management, edited by J. Fraser, B. J. Simkins*, 155–170. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118267080.ch10.

Secchi, D. 2015. “A Case for Agent-Based Models in Organizational Behavior and Team Research.” *Team Performance Management* 21 (1/2): 37–50. doi:10.1108/TPM-12-2014-0063.

Simons, R. 1990. “The Role of Management Control Systems in Creating Competitive Advantage: New Perspectives.” *Accounting, Organizations and Society* 15 (1–2): 127–143. doi:10.1016/0361-3682(90)90018-P.

Sjöberg, L. 2000. “Factors in Risk Perception.” *Risk Analysis* 20 (1): 1–12. doi:10.1111/0272-4332.00001.

Stasser, G., and Z. Birchmeier. 2003. “Group Creativity and Collective Choice.” In *Group Creativity: Innovation through Collaboration*, edited by P. B. Paulus, B. A. Nijstad, 85–109. New York, Oxford: Oxford University Press.

Thagard, P. 1989. “Explanatory Coherence.” *Behavioral and Brain Sciences* 12 (3): 435–467. doi:10.1017/S0140525X00057046.

Thagard, P. 2004. “Causal Inference in Legal Decision Making: Explanatory Coherence vs. Bayesian Networks.” *Applied Artificial Intelligence* 18 (3–4): 231–249. doi:10.1080/08839510490279861.

Thagard, P. 2012. “Cognitive Architectures.” In *The Cambridge Handbook of Cognitive Science*, edited by K. Frankish, W. M. Ramsey, 50–70. Cambridge, UK: Cambridge University Press.

Wall, F., and S. Leitner. 2021. “Agent-Based Computational Economics in Management Accounting Research: Opportunities and Difficulties.” *Journal of Management Accounting Research* 33 (3): 189–212. doi:10.2308/JMAR-19-073.

Wegner, D. M. 1987. "Transactive Memory: A Contemporary Analysis of the Group Mind." In *Theories of Group Behavior. Springer Series in Social Psychology*, edited by B.Mullen, G. R. Goethals, 185–208. New York, NY: Springer.

Appendix A

Table A.1. Overview of the simulation experiments (adapted from Harten, Meyer, and Bellora-Bienengräber, 2022).

	Simulation experiments			
	Limits to information transfer	Incomplete discussions	Group characteristics	Interaction patterns
Experimental conditions				
Receivers retain a part of their prior beliefs	**Yes**[b]	Yes	Yes	Yes
Leader's decision approach	Leader follows own opinion, majority opinion	**Leader follows own opinion, consensus, majority opinion**	Leader follows majority opinion	Leader follows majority opinion
Termination of the discussion	n/a	**At first consensus, after one, five, or ten stable rounds, full discussion**	Full discussion	Full discussion
Differences in the distribution of information	No	No	**Yes/No**	No
Receivers consider hierarchical differences	No	No	**Yes/No**	No
Receivers have no transactive memory	No	No	**Yes/No**	No
Interaction pattern	Random	Random	Random	**Priority to concern vs. dissent vs. hierarchy vs. homogeneity**
Outcome variables	% of correct assessments per discussion round	% of correct assessments, avg. number of discussion rounds	% of correct assessments, avg. number of discussion rounds	% of correct assessments, avg. number of discussion rounds
Number of simulated discussions (n) [a]	1.129	1.129	11.972	11.463
Number of high / low risks	628/601	628/601	5.990/5982	5.701/5.762

[a]Each discussion was simulated until the last piece of information was shared at least once. When deciding on the number of simulation runs, it is necessary to balance computational costs and obtaining representative data generated by the simulation's stochastic process (Lorscheid, Heine, and Meyer, 2012).
[b]Bold highlighting indicates the experimental condition that is the focus of the respective simulation experiment.

Table A.2. Results for simulation experiment 3 (group characteristics) after ten stable rounds.

Differences in the distribution of knowledge	Receivers consider hierarchical differences	Receivers have no transactive memory	Proportion of correct assessments			Avg. number of discussion rounds
			All risks (%)	High risks (%)	Low risks (%)	
+	+	+	65.7%	87.7%	43.3%	14.6
+	+	−	62.4%	81.3%	44.0%	14.9
+	−	+	63.7%	83.7%	44.3%	15.0
+	−	−	64.6%	82.8%	45.3%	14.5
−	+	+	75.1%	86.1%	64.7%	15.7
−	+	−	76.4%	87.2%	65.9%	15.7
−	−	+	75.7%	83.8%	67.4%	15.7
−	−	−	76.4%	87.2%	65.4%	15.7

Note: The table shows the percentage of risks that are correctly assessed after ten stable discussion rounds when the leader follows the majority vote and the average number of rounds before the decision is made, depending on group characteristics that might influence the risk workshop's effectiveness. A '+' indicates the presence of the corresponding deviation from an ideal discussion situation. Differences in the distribution of information are implemented by giving some participants a higher probability of receiving information during the initialization. If receivers consider hierarchical differences, they weigh input according to the sender's hierarchical position compared to their own. With no transactive memory, participants do not distinguish between senders who are experts on the information and those who are not.

Table A.3. Results for simulation experiment 4 (interaction pattern) after ten stable rounds.

Who talks next during the discussion?	Proportion of correct assessments			Avg. number of discussion rounds
	All risks (%)	High risks (%)	Low risks (%)	
Random choice of participants	76.4%	87.2%	65.4%	15.7
Priority to concerned participants	69.9%	93.9%	47.6%	15.3
Priority to participants with dissenting opinions	74.2%	77.8%	70.4%	16.3
Priority to participants with higher hierarchical position	76.0%	86.4%	66.2%	15.6
Priority to participants close to group opinion	73.8%	93.2%	54.8%	14.9

Note: Percentage of risks that are correctly assessed after ten stable discussion rounds when the leader follows the majority vote and the average number of rounds before the decision is made, depending on the interaction pattern. Concerned participants are those who assess the risk as particularly high. Dissenting participants and those close to the group opinion are determined by measuring the distance between their risk assessment and the average risk assessment of the group.